

Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2003 Proceedings

International Conference on Electronic Business
(ICEB)

Winter 12-9-2003

Investment Strategy Analysis using Support Vector Machines

Sheng-Tun Li

Weissor Shiue

Meng-Huah Huang

Shih-Yu Tseng

Follow this and additional works at: <https://aisel.aisnet.org/iceb2003>

This material is brought to you by the International Conference on Electronic Business (ICEB) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICEB 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Investment Strategy Analysis using Support Vector Machines

Sheng-Tun Li¹ Weissor Shiue² Meng-Huah Huang³ Shih-Yu Tseng³

¹Institute of Information Management, National Cheng Kung University, Tainan 701, Taiwan

²Department of Finance, National Kaohsiung First University of Science Technology, Kaohsiung 811, Taiwan

³Department of Information Management, National Kaohsiung First University of Science Technology
stli@mail.ncku.edu.tw

Abstract

Investment strategy is the key point of investors who can make profits or otherwise. Investors always focus on their viewpoints subjectively, which may make them fall into the logic puzzle. The purpose of this paper is to integrate the technical analysis of financial markets with an emerging neural network model, Support Vector Machine (SVM), to solve the problem of investment strategy in Taiwan Futures Market (TAIFEX). The evaluation of investment strategy is the most essential task of investment analysis. However, the evaluation is usually time-consuming and laborious for investment experts. An effective and efficient decision support tool could significantly alleviate his/her burden and improve decision quality. The experimental results from a real-case study demonstrate its salient features of generalization and usability compared with original technical analysis.

Keywords: Investment strategy, technical analysis, three-level prices analysis, support vector machines, neural networks

1. Introduction

While investors distribute their assets among various investments, factors such as individual goals, risk tolerance and horizon will be considered. That is the so-called investment strategy. A successful investment strategy improves investors' chances of profit. However, due to the lack of objectivity, investors usually need indicators to help them adjust investment strategy and determine a sensible planning. Technical analysis (TAs) [12] is one of them, which analyzes price and volume data with charts and graphs to predict future market trends. Most TAs, e.g. MACD, KD, are too complicated and sometimes uncertain, which leads to confusion of investment analyzers, so three-level prices analysis is introduced in this paper. Three-level prices analysis is the extension of candlestick charts, and inspired by Elliott's Wave Theory [6] and Fibonacci Numbers [12]. Meanwhile, neural network techniques are introduced to enhance the performance of three-level prices analysis.

Recently, support vector machines (SVM) have been shown another novel approach to improve the generalization property of neural networks [3], [4]. Since its emergence, SVM has been widely applied to many kinds of field in the past few years, such as text

categorization, handwritten digit recognition and face detection. Most of them are with the engineering and science fields. The major advantages of SVM over other neural network models are better generalization performance, smaller number of parameters setting and capability of handling higher dimension data set.

The remaining sections of this paper are organized as follows. Section 2 describes the problem nature of technical analysis. Section 3 introduces the concept of SVM in solving classification problems. Section 4 exhibits the experiments with SVMs, parameters setting. Section 5 concludes this paper.

2. Technical Analysis

It is well understood that the ultimate purpose of investment is to make profits, so the technical analysis focuses on its profitability. Depending on different approaches used in analyzing the financial trading data, different sets of factors are considered. Technical analysts keep devoting to finding out the knowledge of stock price's movement and many famous theories were appeared, like Elliott Wave Theory, Dow Theory or technical indicators.

Many past research have devoted in data mining for the stock trading strategy, but many of them lack a usefulness knowledge presentation or took they effort to increase the accuracy [15].

Therefore, the investors not only need to understand the forecast solution but also need to realize the knowledge of the decision solution. The greatest contribution of the research which devoted to increase accuracy is focus on how to keep the expert's knowledge into the system. If the case appear again in the future, the system can response the solution very quickly.

2.1. Three-Level Prices Analysis

Three-Level Prices Analysis is one of the technical analyses, which is inspired by Elliott's Wave Theory [6] and Fibonacci Numbers [12]. It consists of three basic indicators, i.e. upper price (UP), middle price (MP), and lower price (LP). Table 1 details these three indicators and other notation conventions of 3-Level Price Analysis.

Table 1. Notation conventions of 3-level prices

UP	upper price $UP(t+1) = L(t) + (H(t) - L(t)) \times 1.382$
MP	middle price $MP(t+1) = (H(t) + L(t)) / 2$
LP	lower price $LP(t+1) = H(t) - (H(t) - L(t)) \times 1.382$
O	opening price of candlestick chart
H	highest price of candlestick chart
L	lowest price of candlestick chart
C	closing price of candlestick chart

The basic definition of three-level prices analysis is shown in table 2.

Table 2. Basic definition of three-level prices analysis

Rule 1	If today's close price is higher than UP then tomorrow's price movement range will be between UP and MP. The highest price do not to predict, and the LP will not be appeared.
Rule 2	If the today's close price is between UP and MP, then tomorrow's price movement range will be between UP and MP, and the LP has low probability appear.
Rule 3	If the today's close price is between LP and MP, then tomorrow's price movement range will be between MP and LP, and the UP has low probability appear.
Rule 4	If the today's close price is lower than UP then tomorrow's price movement range will be between LP and MP. The lowest price does not be predict, and the LP will not be appeared.

Rule1 and Rule2 are generalized to be RuleA; Rule3 and Rule4 are generalized to be RuleB.

Table 3. Generalized definition of three-level prices analysis

Rule A	If today's close price is higher MP then tomorrow's close price will be higher than MP.
Rule B	If today's close price is lower MP then tomorrow's close price will be higher than MP.

3. Theory of SVM

Support Vector Machine, pioneered by Vapnik in 1995, is the state-of-the-art neural network technology based on statistical learning [17], [16]. In recent years, it has drawn over-whelming attentions from diverse research communities thanks to its outstanding performance in solving classification problems. SVMs were originally designed for binary classification, whose objective is to construct an optimal hyperplane that the margin of separation between the negative and positive data set will be maximized. However, in practice, the data set of

interest is usually linear nonseparable. In order to enhance the feasibility of linear separation, one can usually perform a non-linear transformation to the data set into a higher dimensional space, the so-called feature space. Unfortunately, the curse of dimensionality makes the non-linear mapping too difficult to solve. One of feasible approaches is the mechanism of inner-product kernel. The ideas of optimal hyperplane and the kernel-product feature space establish the fundamentals of SVM and will be discussed in the following two sub-sections.

3.1. The Optimal Hyperplane

In order to describe the SVM mechanism clearly, we simplify a general classification problem to be binary and separable. Figure 1 exhibits the basic concept of SVM. There exist uncountable decision functions, i.e. hyperplanes, which can separate the negative and positive data set well, but of which only one has the maximal margin. This indicates that the distance from the closest positive samples to a hyperplane and the distance from the closest negative samples to it will be maximized. The hyperplane with maximal margin is the optimal hyperplane, as shown in Figure 1. The data points located on the dash-lines are support vectors, which satisfy the condition with the equality sign, defined in Eq. (3).

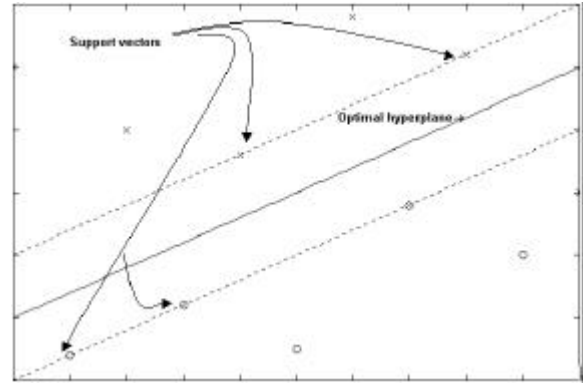


Figure 1. The basis of the support vector machines

Assuming the training data set \mathcal{S} is in this form

$$\mathbf{x}_i \in \mathcal{R}^n, \quad y_i \in \{-1, 1\} \quad (1)$$

$$\mathcal{S} = \{(\mathbf{x}_i, y_i)\}_{i=1 \sim \ell}$$

There are ℓ training examples, where \mathbf{x}_i is n -dimensional input space over \mathcal{R} and y_i is the corresponding output. The decision boundary is defined by the following equation

$$\sum_{j=1}^n x_j w_j + b = 0, \text{ or} \quad (2)$$

$$\mathbf{x} \cdot \mathbf{w} + b = 0$$

Let \mathbf{w} and b be the weight vector and bias of the SVM, respectively. A specific pair (\mathbf{w}, b) represents the one hyperplane. The goal of training a SVM is to find a pair of (\mathbf{w}, b) with the largest margin, subject to the following constraint:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 \quad \text{for } i = 1, 2, \dots, \ell \quad (3)$$

The margin of separation between two classes is $2/\|\mathbf{w}\|$ [7], therefore, the optimization problem is equivalent to minimizing the Euclidean norm of the weight vector \mathbf{w} . The constrained optimization problem, also called the primal problem [3], can be stated as follows:

$$\begin{aligned} &\text{minimize } \Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} \\ &\text{subject to } y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1, i = 1, 2, \dots, \ell \end{aligned} \quad (4)$$

The cost function $\Phi(\mathbf{w})$ is a convex function of \mathbf{w} and the constraints are linear in \mathbf{w} . Based on the context, we exploit the characteristics of Lagrange multipliers [1] to solve the constrained optimization problem, which could be stated as:

$$\begin{aligned} &\text{maximize} \\ &Q(\mathbf{a}) = \sum_{i=1}^{\ell} \mathbf{a}_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \mathbf{a}_i \mathbf{a}_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ &\text{subject to} \\ &\sum_{i=1}^{\ell} \mathbf{a}_i y_i = 0 \quad \text{and} \quad \mathbf{a}_i \geq 0, \quad i = 1, 2, \dots, \ell \end{aligned} \quad (5)$$

where the aiding nonnegative variables \mathbf{a}_i are the Lagrange multipliers, then the weight vector could be stated as

$$\mathbf{w} = \sum_{i=1}^{\ell} \mathbf{a}_i y_i \mathbf{x}_i \quad (6)$$

which only consists of the training data, as well as the objective function $Q(\mathbf{a})$.

Nevertheless, most of classification problems are linear nonseparable, i.e. it is impossible to construct a linear hyperplane without misclassification. The so-called slack variables \mathbf{x}_i are thus introduced to Eq. (3) in order to be tolerant of classification error. In this way, Eq. (3) is redefined as

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \mathbf{x}_i \quad \text{for } i = 1, 2, \dots, \ell \quad (7)$$

That is, the goal of optimization becomes to find out the hyperplane with maximal margin and minimize the probability of classification error.

$$\Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{\ell} \mathbf{x}_i \quad (8)$$

where C is a user-specified parameter which controls the trade-off between complexity of the model and the number of misclassification points.

3.2. Kernel-Induced Feature Space

The introduction of the feature space is to allow the mapping of the input space onto a higher-dimensional space via a nonlinear transformation, which is capable of separating the data set easier.

$$\begin{aligned} \mathbf{j} : \mathbf{x} &\rightarrow \mathbf{x}' \\ \mathbf{x}_i &\rightarrow \mathbf{j}(\mathbf{x}_i) \end{aligned}$$

where \mathbf{x} is the input space, \mathbf{j} is a non-linear transformation and \mathbf{x}_i is the value of \mathbf{x}_i projected to the feature space \mathbf{x}' . Therefore, the cost function of Eq. (5) can be transferred to:

$$\sum_{i=1}^{\ell} \mathbf{a}_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \mathbf{a}_i \mathbf{a}_j y_i y_j \mathbf{j}(\mathbf{x}_i) \cdot \mathbf{j}(\mathbf{x}_j) \quad (9)$$

In order to separate the training data set linearly, one usually projects the input space into an higher-dimensional feature space, which results in the complexity of mapping and the curse of dimensionality. Based on this, the concept of inner-product kernel is introduced as follows.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{j}(\mathbf{x}_i) \cdot \mathbf{j}(\mathbf{x}_j) \quad (10)$$

Instead of calculating the exact value of $\mathbf{j}(\mathbf{x}_i)$, only the inner-product of $\mathbf{j}(\mathbf{x}_i)$ $\mathbf{j}(\mathbf{x}_j)$ is concerned, which is easier to implement. Referred to Eq.(2), Eq.(6) and Eq.(10), the decision boundary is redefined now in the following,

$$\begin{aligned} &\sum_{i=1}^{\ell} \mathbf{a}_i y_i \mathbf{j}(\mathbf{x}_i) \cdot \mathbf{j}(\mathbf{x}) = 0 \\ &\Rightarrow \sum_{i=1}^{\ell} \mathbf{a}_i y_i K(\mathbf{x}_i, \mathbf{x}) = 0 \end{aligned} \quad (11)$$

The kernel function, $K(\mathbf{x}_i, \mathbf{x}_j)$, can be any function which satisfies Mercer's theorem [10], where the radial-basis function is one of the most popular one:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2s^2}\right) \quad (12)$$

Besides, the kernel function can be also polynomial learning networks or two-layer perceptron networks. The determination of appropriate kernel functions is usually case-dependent.

4. Experiments and Evaluation

In this section, we discuss a real case study of applying SVM to evaluating Futures trading strategy in Taiwan.

4.1. Collection of empirical data set

Our experiment data was collected from Taiwan stock exchange, the collected period was between 1991 to 2002, 12 years daily data. Each record includes the Open Price, Highest Price, Lower Price and the Close Price.

The original data is composed of daily candlestick charts; each candlestick chart consists of the opening price, closing price, highest price, and lowest price. Given the candlestick chart one day, the information of the trend next day, especially the closing price will help investors determine your investment strategy. However, the information provided by candlestick chart is limited, so three-level prices analysis is introduced here in order to enhance the consequence. The original data plus three-level prices analysis indicators are like this:

Table 4. Experimental data

Output	O	H	L	C	UP	MP	LP
1	9853	9988	9789	9862	10162	9819	9475
1	9862	9939	9752	9927	10064	9888	9713
0	9927	10053	9879	9964	10011	9846	9681
⋮				⋮			

Because of the nature of SVMs, we try to turn the prediction issue into a simple binary classification problem. The output y of each row is determined by the following equation:

$$y(t) = \text{sgn}(C(t+1) - MP(t+1))$$

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

Output value of 1 means the closing price next day will be higher than the MP generated by current data, and value of 0 means the closing price will be lower. The main purpose of the classification problem is that given a pattern composed of O, H, L, C, UP, MP, and LP, the classifier would determine its corresponding output.

4.2. SVM Simulations

The experiments are conducted by implementing a SVM simulator using OSU SVM Classifier Matlab Toolbox [5] built upon LIBSVM [8], which is capable of selecting parameters of model automatically based on the cross validation accuracy, implements the fast training SMO algorithm [14], and provides a cache memory management mechanism to enhance the operation of the algorithm. The simulation environment is Solaris 7 on SUN Ultra80 with 1GB memory. The Radial-basis function, referred to Eq. (12), is used as the kernel function in order to solve the non-linear classification effectively and the one-against-one algorithm is used to deal with the multi-classification problem, which has been proved to achieve better performance [8], [18].

In order to reduce the bias of simulation, the 5-fold cross validation is conducted. That is the raw data set is randomly permute into five folds; of which fold is used as the validation set and the rest is as the training set. The overall generalization is the averaged generalization on the evaluation set over five runs. The parameter C is determined empirically, which obtains the best performance in generalization. Besides, in order to prove the outstanding performance of SVMs, the experiment is also implemented by radial basis function networks [9]. Radial basis function networks (RBFNs), introduced by Broomhead and Lowe [2], also known as networks with locally-tuned overlapping receptive fields [11], have attracted interest for engineering applications due to their advantages over traditional multiplayer perceptrons in the past decades. Table 4 illustrates the comparison of SVMs with RBFNs and original three-level prices analysis. It is shown that SVM outperforms RBFN.

Table 5. The comparison of SVM and RBFN

	Generalization accuracy	Setting
SVM	67.1836 %	$C = 100, \gamma = 0.001$
RBFN	63.96 %	
3level prices	62.5%	

Table 5 details the accuracy by confusion matrix. The accuracy of targeting class 0, RBFN performs as well as SVM. As to class 1, SVM has a bit higher accuracy.

Table 6. Confusion matrix of SVM and RBFN

Desired \ Computed	0	1
0	66.27%	33.73%
1	35.48%	64.52%

SVM		
Desired \ Computed	0	1
0	66.27%	33.73%
1	36.77%	63.23%

RBFN

Our outcome 67.1836% was calculated by real trading result, we want to know if the accuracy is the limited by the original three-level prices trading strategy or limited by SVM. So we compare the outcomes, we find the SVM is good algorithm, it learned the original very well, the consistency is about 90%. This compare result prove the 67.1836% accuracy is not limited by SVM.

In order to visualize the decision boundary generated by SVM and RBFN, the input space is reduced from 7 indicators to only 2 indicators, C and MP . The visualization is demonstrated in Figure 2.

5. Conclusions

The SVMs has been proved to outperform most other artificial intelligent algorithms in solving classification problem, that's why it becomes so attractive in recent years. This paper is devoted to applying SVMs to investment strategy in which three-level prices analysis introduced. The outstanding performance of SVMs, exhibited in the experimental results, reveals that the SVMs is capable of generalizing well, i.e. avoid overfitting. On the other hand, we also find out that while the training data is chaotic, i.e., the structure in the training data are declared improperly, it will usually lead to lower prediction accuracy. There are rooms for future work. Not only three-level prices analysis, but also various kinds of financial analysis technique could take advantage of the characteristics of SVMs to enhance their performance.

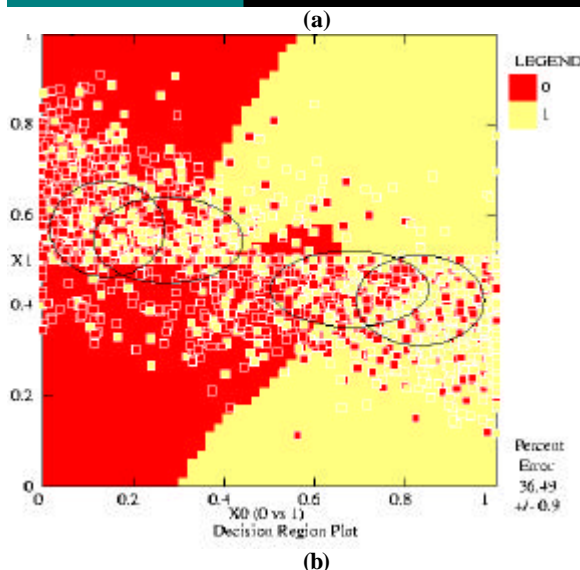
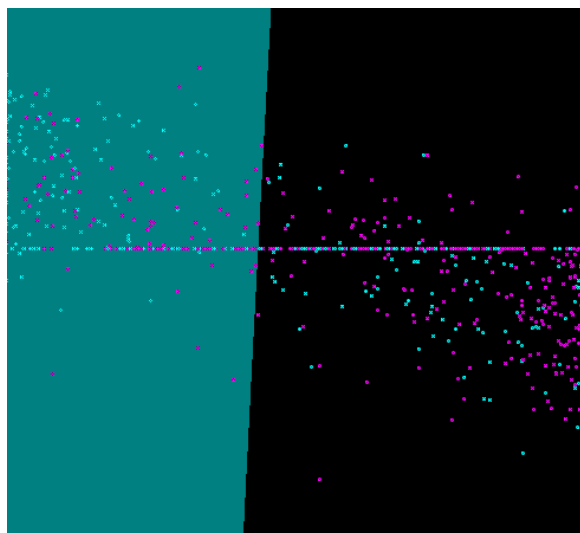


Figure 2. The distribution of decision outcomes by (a) SVM and (b) RBFN

References

- [1] Bertsekas, D. P. "Constrained Optimization and Lagrange Multiplier Methods," New York, Academic Press, 1996.
- [2] Broomhead, D. S., and Lowe, D. "Multivariable Functional Interpolation and Adaptive Networks," Complex Systems, Vol. 2, 1988, pp. 321-355.
- [3] Burges, C. J. C. "A tutorial on Support Vector Machines for pattern recognition," Data Mining and Knowledge Discovery, Vol. 2, No. 2, 1998, pp.121-167.
- [4] Cristianini, N., and Shawe-Taylor, J. "An Introduction to Support Vector Machines and other kernel-based learning methods," Cambridge University Press, 2000.
- [5] Eads, D., Hill, D., Davis, S., Ma, J., Porter, R., Perkins, S., and Theiler, J. "Genetic Algorithms and Support Vector Machines for Time Series Classification," Fifth Conference on the Applications and Science of Neural Networks, Evolutionary Computation, and Fuzzy Systems of the 47th SPIE Annual Meeting, Seattle, WA, 2002.
- [6] Frost, A. J. "Elliott Wave Principle: Key to Market Behavior," John Wiley & Sons, 2001.
- [7] Haykin, S. "Neural Networks: A comprehensive foundation," Prentice Hall, 1999, pp. 318-350.
- [8] Hsu, C.-W., and Lin, C.-J. "A comparison of methods for multi-class support vector machines," IEEE Transactions on Neural Networks, Vol. 13, No. 2, 2002, pp. 415-425.
- [9] Li, S.-T., and Leiss, E.L. "Towards positive-breakdown radial basis function networks," Tools with Artificial Intelligence, Proceedings., Seventh International Conference on, 5-8 Nov 1995, pp. 98 -105.
- [10] Mercer, J. "Functions of positive and negative type, and their connection with the theory of integral equations," Transactions of the London Philosophical Society (A), Vol. 209, 1909, pp. 415-446.
- [11] Moody, J. E., and Darken, C. J. "Fast Learning in Networks of Locally-tuned Processing Units," Neural Computation, Vol. 1, 1989, pp. 281-294.
- [12] Murphy, J. J. "Technical Analysis of the Futures Markets: A Comprehensive Guide to Trading Methods and Applications," Prentice Hall Press, 1986.
- [13] Nison, S. "Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Techniques of the Far East," Prentice Hall Press, 1991.
- [14] Platt, J. C. "Fast training of support vector machines using sequential minimal optimization," Advances in Kernel Methods - Support Vector Learning, Cambridge, MA. MIT Press, 1998.
- [15] Tsai, C.-C., and Wu, S.-J. "A Study of Second-order Modeling of Fuzzy Time Series," in Proceeding of IEEE International Fuzzy Systems in Seoul Korea, August 1999, pp. 719-725.
- [16] Vapnik, V. "Statistical Learning Theory," New York, Springer-Verlag, 1998.
- [17] Vapnik, V. "The Nature of Statistical Learning Theory," New York, Springer-Verlag, 1995.
- [18] Weston, J., and Watkins, C. "Multi-class support vector machines," Proceedings of ESANN99, Brussels, D. Facto Press, 1999.